

Estimation of fuzzy anomalies in Water Distribution Systems

J. Izquierdo^{1a}, M.M. Tung^{2b}, R. Pérez¹, and F.J. Martínez¹

¹Centro Multidisciplinar de Modelación de Fluidos
and ²Instituto de Matemática Multidisciplinar
Universidad Politécnica de Valencia, Spain

Abstract: State estimation is necessary in diagnosing anomalies in Water Demand Systems (WDS). In this paper we present a neural network performing such a task. State estimation is performed by using optimization, which tries to reconcile all the available information. Quantification of the uncertainty of the input data (telemetry measures and demand predictions) can be achieved by means of robust estate estimation. Using a mathematical model of the network, fuzzy estimated states for anomalous states of the network can be obtained. They are used to train a neural network capable of assessing WDS anomalies associated with particular sets of measurements.

Keywords: Water Distribution Systems; Neural Nets; Fuzzy Logic; Modelling

1 Introduction

Water companies use telemetry systems for control and operation purposes. By considering the data provided by telemetry, the engineer on duty makes operation decisions trying to optimize the system utilization. Nevertheless, the system complexity does not permit but to take a few real-time measures, which only incompletely represent the network state. They give indication of only certain aspects of the system, leaving out other more specific or “less relevant” ones. Thus, suitable techniques that allow for more accurate network health estimation are necessary so that anomalies can be detected more rapidly, and light anomalies, which develop progressively and insidiously, can be identified. This will enable to control their consequences in earlier stages, thus avoiding, among other things, losses of water, which can be of great importance.

The state of a WDS is obtained by interrelating different measures within a mathematical model of the network, [Mar95]. Different tools to analyze water networks have been developed in the last years, SARA [GMF98], and EPANET [Ros97], among others. But state estimation cannot be accurately performed if there are missing or uncertain data. Thus, system operators need error limits for the state variables. Yet,

^ajizquier@gmmf.upv.es

^bmtung@imm.upv.es

data are abundant since they are permanently received. Therefore, operators cannot evaluate errors easily or in real time. It is expected that suitable techniques borrowed from Artificial Intelligence (AI) could encapsulate the necessary knowledge to assess the network state.

In this paper, we present an approach for the diagnosis and decision making process which is necessary on a neural network for clustering and pattern classification. First, the mathematical model, a state estimation procedure and a mechanism for treating uncertainties, already presented in [Izq04] and [Izq05], are briefly presented. The state estimator, together with the error limits will be used as a surrogate of the real WDS to generate data to train and check the neural network (NN). Then, the inherent procedures to neural techniques will be described. Specifically, the NN architecture, the classification and clustering mechanisms of both, crisp and fuzzy, patterns and the training technique will be presented.

2 Mathematical Model and State Estimation

Analyzing pressurized water systems is a complex task, especially for big systems. But even for moderately sized cities, it involves solving a big number of non-linear simultaneous equations. The complete set of equations may be written by using block-matrix notation,

$$\begin{pmatrix} A_{11}(q) & A_{12} \\ A_{12}^t & 0 \end{pmatrix} \begin{pmatrix} q \\ H \end{pmatrix} = \begin{pmatrix} -A_{10}H_f \\ Q \end{pmatrix}, \quad (1)$$

where A_{12} is the so-called connectivity matrix describing the way demand nodes are connected through the lines. Its size is $L \times N_p$, N_p being the number of demand nodes and L the number of lines; q is the vector of the flow rates through the lines, H the vector of unknown heads at demand nodes; A_{10} is an $L \times N_f$ matrix, N_f being the number of fixed-head nodes with known head H_f , and Q is the N_p -dimensional vector of demands. Finally, $A_{11}(q)$ is an $L \times L$ diagonal matrix. System (1) is a non-linear problem whose solution is the state vector $x = (q, H)^t$ of the system.

The non-linear relations describing the system balances are complemented by the specific telemetry measurements. These measurements are integrated into the model by expanding system (1) to a new system, typically overdetermined:

$$\begin{pmatrix} A_{11}(q) & A_{12} \\ A_{12}^t & 0 \\ A_{31} & A_{32} \end{pmatrix} \begin{pmatrix} q \\ H \end{pmatrix} = \begin{pmatrix} -A_{10}H_f \\ Q \\ M_t \end{pmatrix}. \quad (2)$$

The components A_{31} and A_{32} in system (2) were introduced to account for additional telemetry measurements M_t with uncertainties in the demand predictions. System (2) is usually solved using least-square methods for a state estimation by an over-relaxation iterative process applied to a linearized version of (2):

$$\begin{pmatrix} A'_{11}(q^{(k)}) & A_{12} \\ A_{12}^t & 0 \\ A_{31} & A_{32} \end{pmatrix} \begin{pmatrix} \Delta q \\ \Delta H \end{pmatrix} = \begin{pmatrix} -A_{10}H_f - A_{11}(q^{(k)})q^{(k)} - A_{12}H^{(k)} \\ Q - A_{21}q^{(k)} \\ M_t - A_{31}q^{(k)} - A_{32}H^{(k)} \end{pmatrix}, \quad (3)$$

where A'_{11} is the Jacobian matrix corresponding to A_{11} .

3 Error Limit Analysis

Error limit analysis is a process to determine uncertainty bounds for the state estimation originated by the lack of precision of measurements and data. To put it in a nutshell, the question is what is the reliability of the estimated state x^* , if measurement vectors y are not crisp but may vary in some region, $[y - \delta y, y + \delta y]$?

Different techniques may be used to estimate this unknown but bounded error, [Mil96], [Nor86], [Kur97]. We use a variant of the so-called sensitivity matrix analysis, [Bar03], which uses the state estimator presented above.

In [Izq05], it is proved that a component by component bound, e^* , for δx^* can be obtained by means of

$$e^* = | (A_k^{*t} W A_k^*)^{-1} A_k^{*t} W | |\delta y|, \quad (4)$$

where W is a diagonal matrix that weights the equations according to the nature of the right-hand sides, and the vertical bars indicate absolute values of all matrix and vector entries. Because of linearity, the bounds calculated by (4) are symmetrical and the error limit may be expressed as a multidimensional interval (see cell definition in next section) $[x^*]$ in the state space

$$[x^*] = [x_{\inf}^*, x_{\sup}^*] = [x^* - e^*, x^* + e^*]. \quad (5)$$

4 The Neural Network

A neural network for clustering and classification is a mechanism for pattern recognition. Here, we use multidimensional cells, [Sim92], [Lik94]. Voronoi diagrams are used in Ref. [Ble97].

A cell C is a region of the pattern space of n -dimensional vectors obtained as the intersection of n pairs of half-spaces of the form $m_i \leq x_i \leq M_i$, for $i = 1, 2, \dots, n$, where m_i and M_i are real numbers. Vectors $m = (m_i, i = 1, \dots, n)$ and $M = (M_i, i = 1, \dots, n)$ are called min and max points of C and completely determine C . Membership of patterns to a cell is defined from fuzzy grounds. For fuzzy patterns, $P = [P^{\inf}, P^{\sup}]$, like the ones obtained in (5), membership values are given by the membership function

$$c(P) = \max_{i=1, \dots, n} \left\{ \max \left\{ \varphi_i (P_i^{\sup} - M_i), \varphi_i (m_i - P_i^{\inf}) \right\} \right\}, \quad (6)$$

where each $\varphi_i(x)$ controls the cell fuzziness.

Values taken by membership function (6) are used during the operation phase to decide the membership degree to the class associated with a cell exhibited by certain pattern presented to it and, as a consequence, to recognize the potential anomalous state of the water distribution system corresponding to the associated label of each class.

Patterns presented to the network during the training phase are ordered pairs (P, l) , where l is a label associated to pattern P describing the type of anomaly it represents.

The NN implementing the classification process is a three-layer network that grows adapting itself to the problem characteristics. The input layer has $2n$ neurons, two for any of the dimensions of the patterns $P = [P^{\text{inf}}, P^{\text{sup}}]$. When a new pattern is presented to the network through the input layer, the components of vectors P^{inf} and P^{sup} are compared, respectively, with those of the minimum point, m , and the maximum point, M , of the J existing cells. Specifically, numbers in the inner brackets of (6) are calculated.

This way, each neuron on the hidden layer has two n -dimensional vectors φ^{inf} and φ^{sup} as its input, formed by numbers between 0 and 1, ready to be processed, first component by component with the max operator, and then with the max operator, but now through all the components. Specifically,

$$c(P) = 1 - \max_{i=1, \dots, n} \left\{ \max \left\{ \varphi_i^{\text{sup}}, \varphi_i^{\text{inf}} \right\} \right\}$$

is calculated for each cell. This process gives the membership degree of P to every one of the cells. Thus, membership functions may be considered as the transfer or activation functions for all the J existing hidden neurons. And the values of the minimum and maximum points of those existing cells, which will be adjusted during the training phase, must be regarded precisely as the synaptic weights between the input and the hidden layer.

The values produced by the membership functions of the existing cells constitute the outputs of the hidden layer. These values must be operated with the weights between the hidden and the output layers. This process will produce a class, a diagnosis of the hydraulic system represented by pattern P . This procedure facilitates the decisions to be made by the system managers.

5 Conclusions

The described neural procedure does not fit into any standard paradigm, since it is made of several sub-nets that evolve by accumulating experience as new loads (peak, valley, seasonal-dependent, etc.) are observed, which mimics human knowledge acquisition.

From the reduced number of tests performed we conclude that the classification ability of the NN is excellent. Since the response given by the NN is graded, as a consequence of its fuzziness, the information it provides is not only qualitative (pointing out an anomaly) but also quantitative (weighting the distributed importance of the problem).

The tool presented here, once completed, calibrated and implemented, will provide WDS managers with a decision support mechanism allowing early identification of anomalies and, as a consequence, better Integrated Water Management.

Acknowledgments. This work has been performed under the support of the projects Investigación Interdisciplinar n° 5706 (UPV), **DPI2004-04430** of the Dirección General de Investigación del Ministerio de Educación y Ciencia (Spain) and FEDER funds.

References

- [Bar03] Bargiela, A., Pedrycz, W.: Granular Computing. Kluwer Academia Press, Boston, Dordrecht, London (2003).
- [Ble97] Blekas, K., Likas, A., Safylopatis, A.: A Fuzzy Neural Network Approach to Classification Based on Proximity Characteristics Patterns. Proc. 9th IEEE Int. Conference on tools with Artificial Intelligence, Newport Beach, CA (1997).
- [GMF98] GMF (Grupo Mecánica de Fluidos): SARA, Software de Análisis de Redes de Agua, Manual de Usuario. Ed. Grupo Mecánica de Fluidos, UPV (1998).
- [Izq04] Izquierdo, J., Iglesias, P.L., Díaz, J.L., Pérez, R.: Identificación de estados en sistemas de distribución de agua mediante técnicas de optimización, IV SEREA Seminario hispano-brasileño: Planificación, proyecto y operación de redes de abastecimiento de agua, CD-ROM, Valencia (2004).
- [Izq05] Izquierdo, J., López, P.A., Pérez, R., Ledesma, B.: Intervalos de confianza en la estimación de estados en sistemas de distribución de agua. V SEREA Seminario hispano-brasileño: Planificación, proyecto y operación de redes de abastecimiento de agua, CD-ROM, Valencia (2005).
- [Kur97] Kurzhanski, A.B., Valyi, I.: Ellipsoidal Calculus for Estimation and Control. Birkhäuser (1997).
- [Lik94] Likas, A., Blekas, K., Safylopatis, A.: Application of the Fuzzy Min-Max Neural Network Classifier to Problems with Continuous and Discrete Attributes. Proc. of IEEE Workshop on Neural Networks for Signal Processing (NNSP'94), 163–170 (1994).
- [Mar95] Martínez, F., R. Pérez and J. Izquierdo: Optimum Design and Reliability in Water Distribution Systems, in Improving efficiency and reliability in water distribution systems. Kluwer Academic Pub., Dordrecht, Boston, London (1995).
- [Mil96] Milanese, M. et al. (Eds.): Bounding Approaches to system identification, Milanese, M. et al. (Eds.), Plenum Press, NY (1996).
- [Nor86] Norton, J.P.: An Introduction to Identification. Academic Press (1986).
- [Ros97] Rossman, L.A.: Manual de usuario de EPANET. Drinking Water Research Group. Risk Reduction Engineering Laboratory. US EPA. Translated by GM Fluidos, UPV (1997).
- [Sim92] Simpson, P.: Fuzzy Min-Max Neural Networks-Part 1: Classification. IEEE Trans. On Neural Networks **3**(5), 776–786 (1992).